AN INTEGRATED APPROACH TO EVALUATING SIMULATION CREDIBILITY

Paul R. Muessig, Ph.D.
Naval Air Warfare Center, Weapons Division (Code 418000D)
1 Administration Circle
China Lake, CA 93555
muessigpr@navair.navy.mil

Dennis R. Laack and John J. Wrobleski
Computer Sciences Corporation
711 Daily Drive
Camarillo, CA 93110
drlaack@csc.com, jwrobles@csc.com

KEY WORDS

Model credibility, model evaluation, verification, validation, accreditation

ABSTRACT

Discussions about simulation credibility tend to focus on the issue of "validation", i.e., how well simulation predictions match real world observations. While validation is certainly a most direct and intuitively accessible measure of simulation credibility, the validation process has numerous well-known limitations. Validation is by no means the only measure of simulation credibility, however. This paper identifies and categorizes a spectrum of information that can be used to evaluate simulation credibility more robustly than reliance on validation results alone. It also describes a method by which the nature, scope and depth of information necessary to establish simulation credibility for a particular application can be determined from an analysis of the risks associated with that application.

INTRODUCTION

What makes a simulation "credible", and what activities contribute to a determination of simulation "credibility"? The modeling and simulation (M&S) community has tended to focus on "validation" activities (i.e., comparisons of simulation outputs with various types of "real world" data) as the primary hallmark of simulation credibility. This emphasis is hardly surprising; it is difficult to imagine a more direct and intuitively appealing measure of simulation credibility. Moreover, the validation process is amenable to scientific rigor; when performed without prejudice in accordance with the scientific method it yields an objective measure of confidence in simulation predictions.

The validation process is not without its drawbacks, however. For example, validation tests tend to be limited in scope relative to the range of predictive capabilities of most simulations, a fact that limits the range over which simulations can be declared "valid". Validation data also tend to be difficult and costly to obtain under conditions that match the predictive constraints of the simulation being evaluated. This means that "real world" data will always contain factors not accounted for by the simulation. Disentangling what is a simulation artifact from what is a data artifact can be a troublesome and speculative process that reduces the value of validation results as a measure of simulation credibility. Finally, some simulations simply cannot be validated in the commonly accepted sense of the term. (Campaign level military simulations come to mind.)

Are there any other measures of simulation credibility? What are those measures, and what kind of confidence do they contribute to simulation outputs? This paper explores one approach to answering these questions, and provides guidance on how to integrate these measures into a robust evaluation of simulation credibility for specific applications.

WHAT MAKES A SIMULATION "CREDIBLE"?

The authors have been involved with all facets of simulation credibility assessment for over ten years. Based on this experience, we have repeatedly seen five key factors that contribute to the evaluation of simulation credibility:

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
00AUG2001	N/A	-

Title and Subtitle An Integrated Approach To Evaluating Simulation Credibility	**Contract Number**
	Grant Number
	Program Element Number
Author(s) Paul R. Muessig, Dennis R. Laack, and John J. Wrobleski	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Naval Air Warfare Center, Weapons Division (Code 418000D) 1 Administration Circle, China Lake, CA 93555	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes	
Abstract see report	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** SAR
Number of Pages 12	

Capability

Simulations are abstractions of reality; they do not simulate all aspects of the "real world", nor do they need to. In order to be considered credible for use in a particular application, a simulation need only represent those aspects of the "real world" that are important to the intended use. These "capability requirements" are derived from an analysis of the application in which the simulation will be used. These requirements must then be compared to actual simulation capabilities to determine whether or not the simulation has all the features necessary to produce credible outputs for the intended use.

Analysis of the intended use of the simulation should lead to a complete description of simulation capability requirements, including:[1]

- A clear description of the intended use of the simulation;
- A listing of the physical entities that must be simulated, the functions they must perform within the simulation, and the degree of fidelity to which these functions must be simulated;
- A description of the environment in which the physical entities will interact, and the rules under which different entities will interact with each other and with the environment.

Descriptions of simulation capability should follow a similar outline:

- A clear description of the purpose for which the simulation was developed;
- A listing of the physical entities included in the simulation, the functions they perform within the simulation, and the degree of fidelity to which these functions are simulated;
- A description of the environment in which the physical entities interact within the simulation, and the rules under which different entities interact with each other and with the environment.
- A summary of assumptions and limitations in simulation design and implementation that impact the scope of potential applications in which the simulation can be credibly used.

Once we know the intended use of a simulation and the capability requirements that flow from it, we are in a position either to specify requirements for simulation development and design or to assess the capabilities of existing simulations against these requirements. The credibility of a simulation is a function of its ability to meet the most important elements of required capability as determined by the intended use. While intuitively obvious in principle, we have observed in practice that both simulation capability requirements and descriptive information about simulations are rarely documented in terms that allow for easy comparison and evaluation.

Software Accuracy

By software accuracy we mean the degree of error-freeness of the simulation software. One must be able to demonstrate on the basis of software test results not only that the software passed all the planned qualification tests, but that the nature scope and depth of those tests was matched to the complexity of the simulation and the risks associated with simulation use. The more complex a simulation is, the more difficult it is to specify and execute a robust software testing program that will unambiguously demonstrate the degree of software accuracy achieved. A poorly scoped software test program with ambiguous or undocumented acceptance criteria can easily result in a relatively meaningless formal "qualification" of the software. In addition, the higher the risks associated with use of the software, the more important it is that the testing be robust enough to demonstrate such accuracy.[2]

Another factor of equal importance to software test results is the quality of the resources applied to software testing. The more complex a simulation is, the more important it is to apply seasoned resources to the planning, management and execution of the test program. It is not enough to rely on "best commercial practices"; one must establish that these practices are actually being applied by qualified personnel. Such an assessment normally requires an evaluation of the software development process used, the quality and appropriateness of the resources applied, and the nature, scope and depth of the artifacts produced.

[1] Readers will recognize elements of sound conceptual model development practice in this description.

[2] One need only refer to the recent disasters with the Ariane rocket system and the Mars lander program to see the impact of even minor software errors on high-risk applications of software.

These factors related to the quality of software test results are frequently overlooked in assessments of simulation credibility, the common practice being instead to treat the results of software testing as a brute fact to be accepted at face value. But software accuracy is a function not only of V&V results; it is a function of how much confidence the simulation user can have in those results.

It is interesting to note that software accuracy will have an important effect on validation efforts as well. If comparisons between test data and simulation predictions do not yield acceptable results, where might the problem lie? The test data could be corrupted or inaccurate; the simulation may not account for some essential feature of the real world; or there may be an error in the software. The more one can demonstrate software accuracy, the easier it becomes to trace unacceptable validation results to data problems or inadequacies in simulation functionality.

Data Accuracy

By data accuracy we mean two things: (1) the appropriateness and error-freeness of all simulation data, and (2) the accuracy of any data transformations performed to convert data from one form to another.

We have found it convenient to divide simulation data into two categories: embedded data and run-time data. Embedded data are those that are typically "hard wired" into simulation software and do not often change. They consist of fixed parameters that characterize a specific physical system or system behavior, as well as physical constants. Examples of the former would be the specific operating frequency and power output of a radar; examples of the latter would be the speed of light and Boltzmann's constant. Run-time data are variable sets of data fed to in to the simulation at run time. They typically consist of data that affect scenario conditions within the simulation. Examples are terrain databases, environmental databases, aerodynamic tables, etc. We make this distinction between types of simulation data because we have found that the rigor with which data accuracy is pursued in simulation development depends in large part on which of these two categories the data fall in.

Embedded data accuracy is normally addressed during software development and testing. We have found, however, that specific documentation of data sources, histories and accuracy assessments tend to be neglected in typical software development documentation. Instead, heavy reliance on the (mostly anecdotal) "corporate memory" of the development effort substitutes for documented evidence of data accuracy. If the development effort is small and the development environment is stable[3] this may pose little risk to the credibility of simulation outputs. For complex simulations developed in more "dynamic" development environments, however, informal and subjective assessments of embedded data accuracy detract from confidence in simulation credibility and make it harder to diagnose errors in simulation output.

We have also observed that the assessment of run-time data accuracy tends to be handled rather passively: as long as the data are obtained from a recognized source, the assumption is made that the data are both appropriate and accurate, often without further inspection. For example, terrain databases obtained from the National Environmental Mapping Agency (NEMA) are automatically assumed to be appropriate, accurate and usable "as-is". Reliance on the credibility of "authoritative data sources" as a guarantee of data accuracy can easily obscure the need for more detailed inspection and assessment of run-time data, especially for high-risk simulation applications.

As noted above, data accuracy is also a function of the accuracy of data transformations within the simulation. It is clear that unit conversions, coordinate transformations, and data pre- or post-processing algorithms all need to be tested to ensure that good data going in do not become corrupted before being acted upon by simulation algorithms. The nature of these activities tends to be informal, however, and their scope and depth need to be assessed against the risk associated with the intended use of the simulation.

While simulation developers and users will agree with all or most of this in principle, we have found in practice that the degree of rigor with which data accuracy is assessed and documented needs to be evaluated in light of simulation complexity and the risks associated with the intended use of the simulation.

Results Accuracy

By results accuracy we mean the degree of correlation between simulation predictions and real world observations. This is where the term "validation" applies, but there are different types of validation, each of

[3] By "stable" we mean either a low turnover rate in the personnel applied to the development effort OR a well documented development audit trail OR both.

which carries implicit assumptions about how the "real world" is defined:

- Validation Against Other Simulations. Comparison of simulation outputs with the outputs of other "accepted" simulations is called "benchmarking". The value of the comparison depends on how credible the "accepted" simulation is. One must be able to document the basis upon which the "accepted" simulation is considered credible in order for the comparison to have any meaning. As with run-time data accuracy, we have observed a tendency to passive acceptance of benchmark simulation suitability.

- Validation Against Expert Opinion. Here, simulation design and outputs resulting from well-defined input conditions are reviewed and evaluated by Subject Matter Experts (SMEs). This process is usually called "face validation". The value of face validation as a contributor to simulation credibility depends upon the nature, scope and depth of SME experience relative to the type of simulation being evaluated. It also depends on the scope and depth of information presented to the SMEs. One must document not only the results of face validation, but also whether or not the right people with the right experience evaluated the right data for the right (i.e., the intended) application.

- Validation Against Test Data. This is what most people think of when the term "validation" is used. It is certainly the most direct and scientifically rigorous means of ascertaining simulation credibility. In addition to the drawbacks mentioned in the Introduction, however, the value of this method of assessing simulation credibility depends on the credibility of the data used to compare with simulation outputs. The validation process should include an explicit assessment of the validity of the validation data. This assessment should include an evaluation of the test instrumentation used, its inherent measurement accuracy, its calibration history, and any other characteristics that might impact the validity of the validation data set. It should also include a description of any data reduction, smoothing or filtering algorithms applied to the data, why these were chosen, and what impact they had on the validity of their representation of the "real world".

Regardless of the type of validation used, the value of the results is proportional to the quality and accuracy of the real world standard against which the simulation is compared. In nearly all cases, actual data from tests or real world observations is considered better than SME judgements, which are, in turn, considered better than benchmarking against another model that represents the real world.

<u>Usability</u>

Our definition of simulation usability is not framed in terms of "ease of simulation use", but rather in terms of "reduced probability of simulation misuse". This operational definition stems from the twin observations that simulations are credible only within a well-defined usage context, and only when they are properly used within that context. Any simulation attribute that reduces the probability of simulation misuse enhances its credibility within a given context.

By usability, then, we refer to that collection of simulation user support features that facilitate credible use of the simulation and reduce the probability that it will be employed inappropriately. Examples of such features are: training in proper simulation use and interpretation of outputs; accurate and comprehensive simulation documentation (User's Manuals, Analyst Manuals, Programmer's Manuals, etc.); on-call technical support for simulation users; simulation user groups that meet on a regular basis; the existence and implementation of a sound configuration management process for the simulation, both during and after development; the availability of trained simulation operators and analysts who can run the simulation and interpret its outputs correctly; and any other support feature that can help simulation users ensure credible use of the simulation.

Of these, configuration management is worthy of special mention as an indicator of simulation usability. Configuration management is the “glue” that ties the version of the simulation being used to all of the V&V results and to simulation documentation. Without a well-structured and effective configuration management program, the simulation user cannot be sure that any of the information used to demonstrate capability, accuracy, or usability really applies to the version of the simulation that is being used. Under these conditions, evaluating simulation credibility will be problematic.

Note that simulation usability is a necessary, but not sufficient, condition for simulation credibility. No simulation user support feature, no matter how well designed to minimize the probability of simulation misuse, will militate against improper use of the simula-

tion. In all cases, the aphorism that "a fool with a tool is still a fool" applies.

HOW MUCH INFORMATION IS ENOUGH?

To determine how much information is enough to establish simulation credibility for a particular application one must know how much credibility the application requires. The level of simulation credibility required is, in turn, dependent on the level of risk associated with how the simulation will be used. The greater the potential adverse consequences of using inappropriate or erroneous simulation outputs, the greater will be the amount of simulation credibility required and the greater will be the amount of information necessary to demonstrate simulation credibility.

Consider the case of two potential applications of an air combat training simulation: training pilots to fly low (i.e., as a standard flight training tool) and training pilots to execute a critical mission to bomb a specific target (i.e., as a mission rehearsal tool). Assume that validation results demonstrate that the simulated radar depiction of a power plant was visually acceptable, but that the geographic location of the power plant was off by 1000 meters.

What is the impact of this result to each of the applications? In the case of the flight training application, the physical location of the power plant is immaterial, and the risk of "negative training" is low. The goal, after all, is to learn to fly low and not hit things, wherever they are. In the case of using the simulation as a mission rehearsal tool, however, the physical location of the power plant relative to surrounding features is critical. In the actual mission the pilot would be expecting the target to be in one location but it would actually be somewhere else, with potentially disastrous consequences for pilot safety and mission success.

Clearly these two cases involve very different levels of risk. Use of the simulation for mission rehearsal requires a greater level of demonstrated simulation credibility than using the simulation for low-level flight training. This means that more information will need to be collected and evaluated to establish simulation credibility for the mission rehearsal application.

Risk is a concept easily understood but difficult to quantify in objective terms. Even so, several texts and documents have attempted to introduce some degree of uniformity in quantifying risk.[4] One very common approach is to consider risk as the product of two components: the impact (or consequences) of an event and the probability or frequency of the event's occurrence. In most cases the factors in this "equation" cannot be quantified absolutely, but can be subjectively evaluated using principles similar to those described in MIL-STD-882C, "System Safety Program Requirements". The general process for determining the overall level of risk first requires quantification of the impact severity and probability for each separately identified risk factor.[5] Using these two elements an overall level of risk is assigned. This process is repeated for each particular risk factor, and the highest level of risk associated with the application is selected as the level that drives the credibility requirement. The criteria used in each step of the process are all subjective, but they are explicitly stated, subject to expert review and consensus, and can be tailored to the specifics of individual problems. The details of the process for determining simulation credibility requirements have been described by Muessig et al.[6]

Once the risks that could accrue from erroneous simulation outputs have been evaluated the amount and type of information needed to make an adequate assessment of simulation credibility can be determined. This is done using a Simulation Credibility Assessment Guide that has been developed by the authors.

SIMULATION CREDIBILITY ASSESSMENT GUIDE

The guide is divided into six major sections (see Tables I through VI). The first section addresses the determination of simulation credibility requirements; the following sections address the various aspects of simulation credibility discussed above. The first column of Tables I through VI identifies the major questions associated with each of the credibility components. The next column of each table identifies the type(s) of information used to answer each of the questions. In many cases, there are several types of information that apply to a single question. The third column identifies specific sources for each information

---

4 See, for example, Steele, Lowell W. 1989. ***Managing Technology, The Strategic View***. McGraw-Hill, pg 118.

5 A risk factor is a specific type of outcome or result. For example, one risk factor might be injury or death of personnel; another might be damage to equipment; a third might be damage to a particular part of the environment.

6 Muessig, P. R., Laack, D. R. and Wrobleski, J. J. "Optimizing the Selection of VV&A Activities A Risk/Benefit Approach." In ***Proceedings of the 1997 Summer Computer Simulation Conference***. Arlington VA. pp 855-860.

type. These three columns basically define the information space that establishes simulation credibility.

The three columns on the far right side of each table provide guidance as to which information source or sources are needed to mitigate particular levels of risk. Note that greater levels of application risk require more detailed information to establish simulation credibility. Note also that the assignment of specific information requirements to specific levels of risk is subjective, and should be tailored to meet the requirements of individual applications. The assignments listed here are reasonable based on the authors' experience. In some cases, the table provides some flexibility to allow the user to select from two or three alternative information sources to establish the required level of credibility.

In practice, the process outlined above, and the credibility assessment tables that result, would work as follows. To make a robust assessment of simulation credibility one needs to know how much credibility the intended use of the simulation requires in the capability, accuracy and usability dimensions, and what information about a simulation's capability, accuracy, and usability characteristics will establish this level of credibility. The risk analysis approach described above would be used to determine how much and what type of information would be required to establish the required level of simulation credibility. These information requirements would then be compared with available information about the simulation, and a list of credibility shortfalls would be compiled. Each element of this list would then be evaluated for its impact on application risk. Unmet requirements for simulation credibility that have acceptable (i.e., low risk) work-arounds would be removed from the list. Unmet requirements for simulation credibility that have no acceptable work-arounds would generate a requirement for more detailed information in the appropriate category.

It should be noted that the Tables provided are only examples of the output of the process described above. Simulation users must tailor the guidance provided here to the particular circumstances of their specific applications. The guidance provided here establishes a clear set of criteria for developing a simulation credibility assessment plan. It also provides a clear basis for explaining the logic that underlies such a plan, thereby permitting an independent assessment of its validity.

SUMMARY

The benefit of categorizing the key elements of simulation credibility as we have is that it provides a convenient way to associate standard V&V activities with the types of credibility they provide. Our categories of simulation credibility also serve to point out areas where standard V&V activities fall short of fully addressing all aspects of simulation credibility, and they suggest other types of information that might be equally important. The result is a more robust set of metrics by which simulation credibility can be evaluated. The risk techniques outlined here then allow simulation users to determine how much and what specific types of information are needed to establish sufficient credibility for their intended application. This information can form a convenient basis both for V&V and simulation credibility assessment planning.

ABOUT THE AUTHORS

<u>Paul R. Muessig</u>

Dr. Muessig is currently Director of the Joint Accreditation Support Activity (JASA) at the Naval Air Warfare Center, Weapons Division, China Lake, CA. JASA's goal is to support weapons system acquisition and testing programs with credible integration of M&S into program objectives. Dr. Muessig has extensive experience in VV&A methodology development, planning and execution in support of numerous M&S applications in acquisition and testing. He has also worked as a defense analyst at the Center for Naval Analyses in Washington, D.C., contributing to survivability analyses for advanced technology aircraft, and leading model validation efforts for the Advanced Low Altitude Radar Model (ALARM). Dr. Muessig holds a B.S. in Chemistry from St. Joseph's University and a doctorate in Physical Chemistry from Brown University.

<u>Dennis R. Laack and John J. Wrobleski</u>

Messrs. Laack and Wrobleski are retired combat Naval Aviators with extensive experience in directing and managing research, development, acquisition, and testing of naval aircraft, weapons and related systems. For the five years they have provided management and technical assistance to the Joint Accreditation Support Activity (JASA), including original contributions in the areas of accreditation requirements determination, M&S acceptance criteria development, and V&V requirements derivation. Mr. Laack holds a B.E.E. from Marquette University, an M.S. in Systems Acquisition Management, and both an M.S. and an Ae.E. in Aeronautical Engineering from the Naval Postgraduate School. Mr. Wrobleski holds a Bachelor of Aerospace Engineering and Mechanics from the University of Minnesota, and both an M.S. and Ae.E. in Aeronautical Engineering from the Naval Postgraduate School. Both are employed by Computer Sciences Corporation in Camarillo, California.

Table I: M&S Requirements Matrix

M&S Credibility Requirements Issue	Items Required	Item Description	Typical Sources	Level of Documentation Needed When Risk Is…		
				Low	Moderate	High
What do you need the simulation to do?	Application description and M&S requirements	The Application Description defines the overall problem to be solved, the outputs needed from the simulation, and how these outputs will be used in its solution. Specifically describes what the model will be used for and how it will be used in the context of the problem. Also specifies required input data sets (e.g., embedded system parameters and constants, look-up tables, run-time inputs, etc.) and data manipulation requirements (coordinate conversions, unit transformations, etc.).	Application Description must be developed for each specific application based on what the simulation will be used for. If the simulation is being developed for a specific application, the Application Description and some of the M&S Requirements can sometimes be derived (or inferred) from top level S/W Design documentation (e.g., A-Specs (Navy), B-Specs (Air Force), etc.) or Analysis Plans. Some programs develop an "intended use statement" that identifies some of the items mentioned, but it is rarely comprehensive enough by itself to meet the requirements of this information element.	Informal documentation (e.g. briefing, memos, etc.)	Formal documentation	Formal documentation with documented management review and approval
How much confidence do you need regarding the accuracy of simulation?	Risk analysis results	Identifies type of risks arising from potential errors in simulation outputs, assesses the probability of the risk actually occurring, and determines the impact of the risk. Result is an assessment of the level of risk associated with the application, which sets the scope and depth of V&V (and related) activities required to mitigate this level of risk.	Must be developed for each specific application. It is essential to obtain consensus on risk elements, impacts, probabilities and levels among technical, operational, and management personnel. These people must be intimately familiar with the decisions to be made based on simulation outputs, as well as the system(s) being simulated and their use. The risk analysis must be based on the intended use defined in the Application Description.	Informal documentation (e.g. briefing slides, memos, etc.)	Formal documentation	Formal documentation with documented management review and approval
What capabilities and expertise are required of the operators and analysts to properly operate the simulation and interpret its results?	Listing of H/W & S/W needed to properly operate the simulation as well as H/W and S/W systems available to users	A list of computer hardware, pre- and post-processors, and system software on which the simulation can run properly; and list of systems and software the user intends to use.	This information can usually gathered or inferred from user documentation, the model manager, or previous users. The user can identify the systems available for simulation operation.	Informal listing of compatible H/W and S/W and available H/W and S/W is acceptable	Informal listing of compatible H/W and S/W and available H/W and S/W is acceptable	A documented listing of compatible H/W and S/W and available H/W and S/W is needed
	Required operator qualifications	Information that identifies the expertise or training that is needed by simulation operators to properly run the simulation and obtain correct and repeatable results.	This information can usually gathered or inferred from user documentation. In some cases it may be necessary to query the model manager or previous users to obtain more detailed requirements.	Informal list of requirements that are identified by the intended operators is acceptable	Informal list of requirements that are identified by the intended operators and reviewed by the next level of management is acceptable	A formally documented list of requirements that are identified by the intended operators and reviewed by the next level of management is needed
	Required analyst qualifications	Information that identifies the expertise or training that is needed by analysts to properly collect and interpret the simulation outputs.	This information can usually gathered or inferred from user documentation. In some cases it may be necessary to query the model manager or previous users to obtain more detailed requirements.	Informal list of requirements that are identified by the intended operators is acceptable	Informal list of requirements that are identified by the intended operators and reviewed by the next level of management is acceptable	A formally documented list of requirements that are identified by the intended operators and reviewed by the next level of management is needed
What degree of User support is required for credible use and interpretation of the simulation?	Description of user support requirements to ensure correct operation by intended operators/analysts	Identifies the support requirements for properly running the simulation. These include any requirements for operator and/or analyst training, user documentation, input databases, on-call support, etc. that would be needed to allow the intended operators and analysts to properly run the simulation and correctly interpret its outputs.	Support requirements to ensure that intended user personnel can properly run the simulation depend on the qualifications and experience of the these personnel. Although there are no definitive means of defining these support requirements, it is obvious that the type and amount of support is inversely related to the experience and tranining of the intended users.	Support requirements defined by intended users and informally documented (e.g. briefing, memos, etc.)	Support requirements defined by intended users and reviewed by higher management level. Informal documentation is acceptable.	Support requirements defined by intended users and reviewed by higher management level. Formal documentation is needed. In addition these requirements will be reviewed and updated based on initial experience in using the simulation

Table II: M&S Capability RequirementsMatrix

<table>
<tr><th rowspan="2">M&S Capability Issue</th><th rowspan="2">Items Required</th><th rowspan="2">Item Description</th><th rowspan="2">Typical Sources</th><th colspan="3">Type, Scope and Depth of Information Required When Risk Is…</th></tr>
<tr><th>Low</th><th>Moderate</th><th>High</th></tr>
<tr><td rowspan="10">Does the simulation do what you need it to do?</td><td rowspan="4">Functional breakdown and description of simulation</td><td rowspan="4">Describes what the simulation actually does. Must describe: simulation purpose; functions/objects included and the relationships between them; the level of fidelity at which each function or object is represented (e.g., algorithm descriptions, decision logic, lookup tables, etc.); function/object level I/O and I/O relationships between them.</td><td>User documentation (Users' Manual, Programmers' Manual, Analysts' Manual, Version Description Document, training manuals)</td><td rowspan="4">Required. General description of simulation capability at the level of detail typically found in a User's Manual is sufficient.</td><td rowspan="4">Required. Documented description of simulation and functional/object breakdown at a level of detail typically found in an Analyst Manual. Must be sufficient to evaluate simulation and/or object fidelity against the requirements of the application.</td><td rowspan="4">Required. Formally documented description of simulation and function/object breakdown at level of detail typically found in a Conceptual Model Description. Must be sufficient to evaluate simulation and/or object fidelity against requirements of the application.</td></tr>
<tr><td>Software design documentation, possibly including Data Flow Diagrams and source code</td></tr>
<tr><td>Conceptual Model documentation.</td></tr>
<tr><td>Previous Accreditation Support Packages</td></tr>
<tr><td rowspan="6">Summary of assumptions, limitations and errors</td><td rowspan="6">Describes assumptions, limitations and known errors that are implicit or explicit in the model's design and/or coding, and summarizes their impact on simulation use. Where appropriate, impacts should be correlated with each of the functions in the Functional Breakdown as well as the overall simulation level.</td><td>Software design documentation</td><td rowspan="6">Required. Assumptions, limitations, errors and impacts at the simulation level should be identified and cross-referenced to existing current documentation.</td><td rowspan="6">Required. Assumptions, limitations, errors and impacts at the simulation and function/object levels should be identified and cross-referenced to existing current documentation.</td><td rowspan="6">Required. Assumptions, limitations, errors and impacts at the simulation and function/object levels should be consolidated from existing current documentation and formally documented.</td></tr>
<tr><td>User documentation (Users' Manual, Programmers' Manual, Analysts' Manual, Version Description Document, training manuals)</td></tr>
<tr><td>Configuration management databases are useful sources for known errors.</td></tr>
<tr><td>Some assumptions and limitations may also be found in verification or validation reports but may not be explicitly stated as an assumption, limitation, or error.</td></tr>
<tr><td>The Version Description Document may be useful source of information on known errors not addressed in the current version, as well as possible workarounds.</td></tr>
<tr><td>Previous Accreditation Support Packages are also a valuable source of information.</td></tr>
</table>

Table III: M&S Software Accuracy Requirements Matrix

M&S S/W Accuracy Issue	Items Required	Item Description	Typical Sources	Type, Scope and Depth of Information Required When Risk Is…		
				Low	Moderate	High
How much confidence do you have in the accuracy of the software?	S/W development and maintenance process description	The Simulation Development Process description should cover the entire simulation life cycle, from initial development to operation and maintenance. It should include a description of the development paradigm and its implementation; a description of any software development and management tools used; a logical process for defining, tracing, testing and documenting requirements throughout software development and maintenance; configuration management covering the entire simulation life cycle; and adequate provision for documentation of all of these activities. Processes should also exist for keeping all documentation consistent and current with the software.	Look for a S/W Development Plan (SDP), S/W Management Plan (SMP) or a Configuration Management Plan. If these documents are unavailable, look for other documentation that describes the life-cycle management activities.	A top level process description is required. Description should address process for defining and tracing requirements, S/W development and testing, and configuration management	A top level process description is required. It should address all issues for low risk applications as well as the development paradigm, how V&V activities are integrated with development, and processes for updating and regression testing of the software.	A formally documented and detailed process description is required.
	S/W development and management resources description	The resource description should include a description of the H/W environment and the S/W engineering tools that will be/were used; the qualifications of the personnel who will/did code the S/W and perform CM functions; and who will be/was responsible for production of key documentation and testing. A history of the developer's experience with simulation development should also be included.	Check the SDP or other management plans that might contain such information. If this information isn't in existing documentation, discussion with the software developers and managers will be necessary to obtain as much of this information as possible, even if anecdotal. Evidence of simulation development qualifications may be available in SEI Capability Maturity Model evaluation reports.	Not required.	Desirable.	Required.
	S/W development and management artifacts and documentation	"Artifacts" refers to the evidence that S/W development and management are actually being implemented in accordance with the process described in the documents identified in row 1. Such artifacts are usually informal in nature and not deliverable items. Documentation consists of deliverable items from the development effort, and must comply with known (or acceptable) standards and practices for format, content, currency and applicability to the current version of the S/W.	Look for standard documentation that indicates that a disciplined software development and management process was/is being followed. The most important examples are configuration management histories and logs, current model documentation (User Manual, Programmers' Manual, etc.), S/W design documentation (particularly a documented set of requirements and a conceptual model), requirements trace reports, reports of design reviews, peer reviews, and/or logical reviews, code walk-through reports, and S/W Problem Change Request logs, configuration management database status reports, System Change Requests (SCRs) and/or System Trouble Reports (STRs), CCB and User Group meeting minutes.	Required. Number, scope and depth of artifacts should be commensurate with process description above.	Required. Number, scope and depth of artifacts should be commensurate with process description above.	Required. Number, scope and depth of artifacts should be commensurate with process description above.
	S/W verification results	Includes all evidence that the code has been developed according to the design and is free of critical errors. The types of results will include reports from design reviews, code walk-throughs, regression testing on model changes, S/W testing, and supplemental V&V efforts of previous M&S users.	Module, subsystem and system S/W test reports; S/W Problem Change Request (SPCR) logs that correlate verification results with specific versions of the S/W; alpha- or beta- test reports for both new requirements testing and regression testing; specific verification reports for the M&S version being used. IV&V reports may also be useful.	System level verification test results desirable.	System and subsystem level verification test documentation is required.	System, subsystem and module level verification test documentation is required. IV&V results are desirable.
	S/W Quality Assessment	A formal assessment, by someone independent of the software developer/manager, of the complexity, programming conventions, and other indicators of software quality.	This assessment is an independently performed task that is normally reported in a formal document. It may be available through the model manager or an M&S repository.	If formally documented, may be substituted for the S/W development and management process description identified in the first row.	If formally documented, may be substituted for the S/W Development and Management Process description identified in the first row.	If formally documented, may be substituted for the module level test documentation identified in the row above.

Table IV: M&S Data Accuracy Requirements Matrix

M&S Data Accuracy Issue	Items Required	Item Description	Typical Sources	Needed When Risk Is…		
				Low	Moderate	High
How much confidence do you have in the quality and suitability of input data obtained from outside sources?	Indications of data quality	Information that establishes the quality of the data in a database. Typically this information consists of a body of metadata that describes the database, its source, specifications, intended usage, history, and how it was collected. Metadata elements primarily exist at the database level, however, some information at the data element level is sometimes required. This metadata might be provided by the data producer or generated by the user.	Metadata elements generated by the data producer should be available in the same archives that contain the database itself. A list of useful metadata elements that are useful for accreditation is contained in attachment 1 Excel Workbook. If the metadata indicating database quality is inadequate, incomplete, or missing, the user must assess the quality of the database. For databases generated from tests, information similar to that described in Attachment 1 can generally be found in documents such as test plans, laboratory procedures, calibration records, test reports, etc. that governed the development of the database. For databases generated through surveys or observations of operations or real world interactions, information that indicates the quality of this data can generally be found in data collection plans, reports, and raw notes.	Scope and depth of Information needed is indicated in Attachment 1	Scope and depth of Information needed is indicated in Attachment 1	Scope and depth of Information needed is indicated in Attachment 1
How much confidence do you have in the quality and suitability of self-generated input data?	Indications quality assurance in the data generation process	An assessment of the process, equipment, tools, instrumentation, etc. used in generating the data. This assessment should generate information similar to that included in the critical metadata elements of the Data Quality Profile.	Information that indicates the quality of data collection procedures can generally be found in documents such as test plans, laboratory procedures, calibration records, test reports, etc. Information that indicates the quality of data collected through surveys or monitoring operations can generally be found in data collection plans, reports, and raw notes.	An assessment of the quality and suitability of the data collection process is required. Informal documentation of this assessment is acceptable.	An assessment of the quality and suitability of the data collection process and the resources used in the process is required. Informal documentation of this assessment is acceptable.	An assessment of the quality and suitability of the data collection process and the resources used in the process is required. Formal documentation of this assessment is required.
How much confidence do you have in the accuracy of the data manipulations?	Indications of data manipulation accuracy	"Data manipulation" includes user operations on the data such as: editing, subset selection, merging, aggregation, transformation, estimation, interpolation, etc. Indications of manipulation accuracy includes independent reviews of these manipulations, checks or comparisons of the data before and after the manipulation, testing the manipulation process with known sets of data, or any other activities that prallel software verification activities.	Verfication of data manipulation procedures may be documented in M&S verification reports (when done in conjunction with M&S development). Other data manipulation should be reviewed and verified as part of the M&S accreditaiton process and documented in the accreditation report. This documentation should describe the verification techniques that were used.	Required at cursory level. Informal documentation of the verification steps is acceptable.	Required. At a minimum the verification should include an independent review (may be informally documented) or a formally documented verification process followed in conjunction with M&S verification.	Required. At least two separate, formally documented verification steps should be included.

Table V: M&S Output Accuracy Requirements Matrix

M&S Output Accuracy Issue	Items Required	Item Description	Typical Sources	Type, Scope and Depth of Information Required When Risk Is…		
				Low	Moderate	High
How much confidence do you have in the simulation outputs?	Benchmarking Results	Documents the results of comparisons between simulation (or simulation component) outputs and those of a "standard" or widely accepted simulation. Benchmark simulations are generally of greater fidelity than the simulation (or component) under review and are characterized by some "stamp of approval" from a recognized authority. Benchmark results should include the name and source of the standard simulation, why it is (or should be) considered a "reference" simulation, which parameters between simulations (or components) were compared (and why), what the results of the comparison were, and what these results imply about the credibility of the outputs from the simulation under review.	Benchmarking results are usually documented in either a validation report, a briefing that describes the results of the comparison, or an accreditation support package. These reports would generally be prepared by previous users. They might also be available through the model manager or developer or DoD M&S databases (e.g., MSRR).	Evidence of any completed validation is required. Documentation can be informal.	System level and appropriate function / object level validation is required from either face validation or results validation, with formally documented results.	System level and appropriate function / object level validation is required from at least two validation techniques, with formally documented results. Whenever possible, results validation should be included as one of the two validation techniques.
	Face Validation Results	Describes the results of subject matter expert opinions about simulation realism and accuracy. This should be based on a structured review of simulation (or component) outputs, sensitivities, and may also include a review of simulation design. Documentation should describe which aspects of the simulation were reviewed (and why), who participated in the review, why one should trust their opinions (e.g., biographical information), what the results of the review were, and what these results imply about the credibility of the simulation. When face validation includes a review of the simulation design, the documentation should also state whether the representations are realistic and whether any assumptions that underly the design are acceptable from the perspective of the intended use.	Face validation results are typically documented in a face validation report (or an accreditation support package) or a previous accreditation assessment report (if the face validation was done as part of an accreditation assessment). If the review was a validation of the design, the results may be reported in a design review report (either a formal report or an annotated briefing). These reports would generally be prepared by previous users. They might also be available through the model manager or developer or DoD M&S databases (e.g., MSRR).			
	"Results" Validation Documentation	Describes the results of comparisons between simulation (or component) outputs and data collected from tests, exercises or operations involving the real system(s) or process(es) being simulated. The documentation should include a description of the source data used in the comparison, from where and how it was obtained, and why it should be considered representative of the real world. Issues relating to data quality (e.g. instrumentation accuracy, calibration, test scenario realism, etc.) should be addressed in the validation report. The correlation between model outputs and real world data should be stated in quantitative terms rather than merely stating that the correlation was "good". Statistical methods should be described and justified. Any anomalies and their impact on model usage should be explained.	"Results" validation is typically documented in a validation report (or an accreditation support package). In some cases, results validation might be documented with an annotated briefing. These reports would generally be prepared by the simulation developer or previous users. They might also be available through the model manager or developer or DoD M&S databases (e.g., MSRR).			
	Sensitivity Analysis Results	Describes the results of experiments to determine the variation in simulation outputs for various measured changes in inputs, functional operations, or other conditions. Sensitivity analyses may be done at the overall simulation level, subsystem level, or module level.	Sensitivity analysis results are typically documented in a separate report. In some cases such results might also be documented in a validation or verification report if the sensitivity analysis was done as part of and contributed to a more comprehensive verification or validation process.	Not required	These results may be used in conjunction with another validation method to reduce the scope of the validation with the other method.	These formally documented results may be used to reduce the scope of the required validation with the other two methods.

Table VI: M&S Usability Requirements Matrix

M&S Usability Issue	Items Required	Item Description	Typical Sources	Type, Scope and Depth of Information Required When Risk Is…		
				Low	Moderate	High
Can you run the simulation properly and interpret the outputs credibly?	Demonstration of the computer hardware and operating system suitability	Test results that show that the hardware and operating systems used to host the model or simulation (if different than that used to develop the M&S) will allow it to run correctly and produce consistent results across platforms.	Information on M&S portability across platforms is usually found in the user documentation associated with the simulation. If this information is not documented, test results will be needed to demonstrate portability.	Informal evidence required.	Documented evidence required.	Documented evidence required.
	Evidence of proper interface and operation of pre- and post-processors	Information that shows that any auxiliary tools and utilities used to format or load input data, or to convert, record, and visualize model outputs are properly interfaced with the simulation being used and operate properly.	User documentation associated with the simulation may list tools and utilities that are utilized or are compatible with it. If this information is lacking, user documentation for the tools and utilities may contain information that will aid the determination of tool compatibilty with the simulation. For non-COTS/GOTS tools and utilities, verification of proper interface and operation is the responsibility of the user.	Informal evidence required.	Documented evidence required.	Documented evidence required.
	Operator qualifications	Information to demonstrate that the operators running the simulation have the expertise and knowledge to properly set up the simulation, execute it, and operate all associated tools and utilities. Typical information includes experience with the specific model being used, formal training on the model, and experience with the hardware, software, and interface devices being used.	This information is usually gathered from biographies or interviews with the operators.	Informal evidence of ability to run the simulation required.	Documented evidence of experience in running the simulation in one or more previous applications required.	Documented evidence of extensive experience in running the simulation.
	Analyst qualifications	Information to demonstrate that the analysts using the simulation have the expertise and knowledge to properly generate the input data and interpret the outputs. Typical information includes experience with the specific model being used, formal training on the model, experience in performing similar analyses and experience or training in M&S based analysis techniques.	This information is usually gathered from biographies or interviews with the analysts. It may also be found in prior accreditation assessment reports.	Informal evidence of basic analytical skills in similar or related applications is required.	Documented evidence of experience in use of the simulation for similar applications.	Documented evidence of extensive experience in use of the simulation for similar applications.
	Availability of user support services	Documentation, training and other user support (e.g., on-call technical assistance, web sites, user groups, etc.) available to establish and maintain user profiency and qualifications in simulation operation and interpretation of outputs.	Documented or summarized in such sources as DoD M&S databases and repositories. Model managers, user groups and user documentation are also good sources.	User documentation is required. User groups and training are desirable.	User documentation is required. User groups and training are desirable.	User documentation is required. User groups and training are desirable.
Does the simulation have community acceptance?	Usage / accreditation history	Summary of prior uses of the simulation, including description of the application and its scenario conditions, which version of the simulation was used, who used it, whether or not it was formally accredited (and by whom). Any documented problems or issues associated with simulation use for the application should also be included.	Accreditation support packages, generic DoD or service M&S study reports, user group meeting minutes and other documentation, model manager, model web sites, M&S databases and repositories, and final reports generated for specific applications. Inclusion in service or M&S databases and repositories (e.g., SURVIAC, Air Force Standard Analysis Toolkit, etc.) is also a valuable indicator of community acceptance.	Not required.	Desirable.	Required.